



COLLOQUE ASL 2009

(5 décembre 2009 – Sorbonne – Amphi Guizot)

SCIENCES DU LANGAGE ET NOUVELLES TECHNOLOGIES

RÉSUMÉS DES INTERVENTIONS

1. NORMALISATION

Benoît HABERT

« *Normalisation, catégorisation, transmission* »

La mise en place d'instruments dans une discipline, pour mesurer ou pour agir sur les objets du domaine, implique des normalisations, c'est-à-dire de s'accorder, pendant une période donnée, sur un « état des choses » et sur des manières d'interagir avec lui. La transposition en sciences du langage, au motif de l'irruption de l'ingénierie linguistique, ne va pas de soi : les « êtres » linguistiques relèvent-ils d'une telle approche ? On utilisera les travaux d'Alain Desrosières pour contraster les manières de mesurer dans différents domaines et pour les relier à la question de la catégorisation et de la stabilisation des catégories utilisées en sciences humaines et sociales. S'en trouvera partiellement transformée la question de la transmission au long cours des résultats des recherches en sciences du langage.

Kim GERDES

« *Les ressources linguistique normalisées et ouvertes : définitions et importance* »

Avec la montée des logiciels libres, nous assistons à une prise de conscience dans la communauté linguistique : le partage des idées, un pilier de la science, inclut aussi les ressources sur lesquelles on travaille, surtout parce que seules les ressources permettent de vérifier ou falsifier les affirmations publiées et que la reconstitution de ressources, déjà existantes ailleurs, prend inutilement une partie importante des budgets limités de la recherche publique et privée. Pourtant, pour l'instant, ces ressources libres restent extrêmement limitées, surtout pour des langues autres que l'anglais.

Nous examinerons à l'aide de quelques exemples les raisons pour cette réticence à libérer ses données :

On rencontrera les contraintes liées à la montée en puissance du droit d'auteur ainsi que des contraintes dans des coopérations industrielles ou plus généralement dans la commercialisation de la recherche. De plus, on se heurte à la méconnaissance de ce que constitue une ressource vraiment libre.

Mais il existe aussi des raisons particulières à la communauté linguistique : une quelconque normalisation présuppose un accord sur l'objet de l'étude ainsi que sur la terminologie. Or il semble que la linguistique en France n'a pas de courant dominant qui pourrait imposer de tels choix.

Nous terminerons sur la description de quelques exemples de ressources libres et des normes qu'elles ont pu instaurer.

Ivan SMILAUER

« *Description formelle et diagnostic automatique des erreurs en L2 : quelques perspectives pour les outils d'enseignement des langues assisté par ordinateur* »

Nous présenterons les possibilités des nouvelles technologies et de la linguistique computationnelle appliquées aux outils d'enseignement des langues assisté par ordinateur. Sur l'exemple de la plateforme www.cetlef.fr, nous montrons que le développement d'un diagnostic automatique des erreurs met en jeu les deux aspects majeurs d'une normalisation des données linguistiques : d'une part, la normalisation du métalangage descriptif et classificatoire des données exigé par le traitement informatique ; d'autre part, la normalisation d'un texte mal formé par rapport à une norme établie. Le diagnostic présenté permet de générer une réponse interactive aux activités langagières des apprenants. En enrichissant une base de données, les productions analysées servent à la recherche sur l'acquisition d'une langue étrangère. Les résultats obtenus dans un cadre expérimental ciblé sur des exercices de flexion nominale tchèque offrent un bon point de départ pour la prise en compte des niveaux linguistiques supérieurs.

2. Analyse du discours

Virginie LETHIER et Jean-Marie VIPREY

« *Nouvelles matérialités textuelles pour l'analyse de discours* »

Nous tenterons d'articuler deux lignes d'intervention. Tout d'abord, un compte rendu succinct de l'expérience de constitution d'une base textuelle numérisée de grande dimension (un journal quotidien régional de la III^e République, de 1883 à 1944 - 200 millions de mots à terme) et d'une recherche approfondie (thèse sur allocation de recherche) à partir d'un corpus extrait de cette base, conduisant à quelques réflexions synthétiques d'ordre théorique et méthodologique. Ensuite, la présentation de quelques éléments plus spécifiquement statistiques et logiciels, en rapport avec l'étude du vocabulaire en corpus, dans une perspective d'essence distributionnelle : en somme, en quoi les « nouvelles technologies » font-elles plus que faciliter les approches même modernes, mais induisent un bouleversement de la matérialité textuelle des discours.

Pascal MARCHAND

« *L'analyse de discours assistée par ordinateur au service d'une théorie de l'usage du langage : une application au discours politique* »

Le concept d'*attitude à l'égard du langage* repose sur la distinction saussurienne « paradigme » / « syntagme », telle qu'elle a été reprise par Jakobson et Halle (1956), puis élargie par Beauvois et Ghiglione (1981). Les opérations de choix (rapport paradigmatique) et de combinaison (rapport syntagmatique) sont envisagées comme autonomes et pouvant faire l'objet d'une centration spécifique sur l'une ou sur l'autre : « en manipulant ces deux types de connexions (similarité et contiguïté)... un individu révèle son style personnel, ses goûts et préférences verbales ». Au-delà, la « prédominance » de l'un ou l'autre de ces deux pôles peut également traduire « certains styles, habitudes personnelles, modes courantes » (Jakobson et Halle, p. 62-63, cité par Beauvois & Ghiglione, p. 77). Un lien peut donc être établi entre une attitude préalable, une activité langagière et des indices discursifs.

Plus précisément, Beauvois et Ghiglione proposent de repérer la directionnalité textuelle (ou investissement dans l'énoncé de tel ou tel registre d'attitude) à partir d'indices langagiers :

– la *directionnalité textuelle* est donnée par les parties de discours qui, soit trouvent leur référence dans le texte lui-même, soit constituent les éléments de l'articulation discursive (« il » locutoire ; pronoms et adjectifs démonstratifs ; relatifs, subordinations et coordinations...);

– la *directionnalité interlocutoire* est donnée par les parties de discours qui traduisent l'intervention directe dans le discours de l'un et/ou l'autre des deux pôles de l'interlocution (pronoms personnels de la première et deuxième personne du singulier et du pluriel ; pronoms et adjectifs possessifs ; les séquences ego- et sociocentriques).

L'objectif sera ici de tester ce modèle au moyen de logiciels d'analyse de discours assistée par ordinateur (lexicométrie, analyse morphosyntaxique) et sur une situation réelle de communication politique : les déclarations prononcées par dix candidats à l'élection présidentielle de 2007 lors de la signature du « pacte écologique » de Nicolas Hulot (31 janvier 2007). À l'issue de ces analyses, la dimension *attitudinale paradigmatique / syntagmatique* apparaît comme structurant fortement les discours. L'application des nouvelles technologies à une théorie issue des sciences du langage montre ici sa pertinence.

Michael BAKER

« *Écriture et réécriture graphico-textuelle dans les débats pédagogiques médiatisés par ordinateur* »

Cette recherche aborde l'étude des débats à finalité pédagogique sur des questions socialement vives, menés par des lycéens qui interagissent à travers des outils informatiques fonctionnant sur Internet/Intranet. Ces outils, conçus et réalisés dans le cadre de deux projets européens (Programme IST de l'UE), permettent la mise en commun, l'organisation, la réécriture et l'élaboration d'arguments, sous les formes d'un schéma argumentatif partagé interactif et d'un texte collectif. Je présenterai les résultats d'une recherche longitudinale portant sur les usages de ces outils, qui met en évidence les processus de réécriture de l'espace du débat, et les processus d'appropriation des outils. En guise de conclusion, j'aborderai la question de l'évaluation de la qualité du débat, sous ses différentes manifestations sémiotiques.

3. Nouvelles technologies, formes langagières

Sonia BRANCA-ROSOFF

« Ce que changent les nouvelles technologies aux problèmes de constitution et d'exploitation de corpus oraux »

En prenant comme exemple le corpus de français parlé parisien constitué à Paris 3 autour de l'équipe SYLED (CFPP2000 - <http://ed268.univ-paris3.fr/CFPP2000>), j'évoquerai dans une première partie les possibilités offertes par le développement d'Internet et par les logiciels libres comme Transcriber, Audacity, Lexico, etc. Ces outils facilitent considérablement la transcription, le stockage, le partage et la consultation des données et devraient rapidement entraîner l'extension d'une linguistique de l'oral. J'aborderai cependant pour finir quelques difficultés et quelques précautions à prendre dans l'interprétation des résultats.

Dominique CAUBET

« Nouvelles technologies, nouvelles formes langagières : blogs, sms, textos, un pas décisif pour l'arabe marocain (darija) dans les années 2000 »

Grâce aux nouvelles technologies, l'arabe marocain vient de connaître un passage à l'écrit inédit ; il s'agit au départ d'un écrit sur un clavier de téléphone portable ou d'ordinateur, différent de celui réalisé avec un stylo sur papier imprimé. En analysant finement ces écrits qui se comptent en millions de pages quotidiennes dans des cybercafés ou sur des écrans de portables, on peut se poser la question de la limite au niveau des contenus et des auteurs : écriture pratique, communication assez basique, nouvelle scène musicale, textes de chansons, textes comiques, phénomènes urbains, de jeunesse... la littérature n'est pas encore répandue sur la toile, mais on ne peut pas prédire si elle y aura accès dans les années ou les mois à venir. Des sites comme Facebook ont depuis moins de deux ans également bouleversé les pratiques d'écriture : statut, quizz, etc. Ce passage à l'écrit s'est produit en graphie latine pour des raisons pratiques plus qu'idéologiques, et des habitudes graphiques que l'on ne peut sans doute pas encore qualifier d'orthographe se sont mises en place. On note des réticences pour le passage à l'impression de cette écriture ; la presse fait les premiers pas, ainsi que les livrets des CDs des groupes marocains. Nul ne peut prédire l'évolution de ce phénomène que personne n'aurait pu prévoir il y a dix ans.

Isabelle PIEROZAK

« Nouvelles technologies, nouvelles formes langagières : quels enjeux ? »

Cette intervention, en abordant l'une des thématiques du colloque par la problématique des enjeux, propose une réflexion sur le sens que l'on peut donner à « nouveauté » dans ce contexte. S'agit-il encore seulement de technologies et faits linguistiques « nouveaux ? Peut-on parler de « nouveauté » dans la manière d'en rendre compte, autrement dit au plan scientifique ? À quelles conditions ? Il sera plus particulièrement envisagé ici une réflexion d'origine sociolinguistique et « trans-technologies » à partir d'études diversifiées.

4. Parole et nouvelles technologies

Christian Hudelot

« Usage de l'ordinateur comme moyen de « communication facilitée » auprès d'enfants autistes ? »

Je me propose non pas de traiter de ce qu'apportent les nouvelles technologies au traitement des données linguistiques, mais de revenir sur une pratique interlocutive issue d'une « nouvelle technologie ».

Il s'agira de rendre compte, du point de vue d'une linguistique du discours, des énoncés écrits par des enfants autistes privés de langage oral par le biais de la technique dite de « Communication Facilitée » (CF). La CF consiste à maintenir le poignet, le coude ou l'avant-bras de l'enfant assis devant le clavier d'un ordinateur et de l'inciter à écrire. À la suite des résultats spectaculaires de cette méthode, une controverse, parfois violente, s'est développée, mettant en cause le fait que l'enfant soit le véritable producteur des textes ainsi produits.

Après un rapide bilan des travaux de psychologie expérimentale qu'a entraînés cette méthode, je rappellerai que la question de savoir qui est l'énonciateur des énoncés ainsi produits se pose sans doute différemment en sciences du langage et en psychologie cognitive. Je me propose enfin, à partir de quelques exemples, de montrer comment le dispositif produit un discours argumentativement orienté.

Christophe PARISSE

« *Corpus multimédias et recherches sur l'acquisition du langage* »

Depuis plus de vingt ans, de nombreux chercheurs ont constitué des corpus d'acquisition de langage destinés à être partagés par l'ensemble de la communauté scientifique. Aujourd'hui, les corpus de textes sont de plus en plus souvent augmentés de données sonores ou visuelles également partageables. De manière concomitante, de nombreux outils existent et sont développés pour accéder à ce type de données et les utiliser dans le travail de recherche.

Il en résulte aujourd'hui une modification progressive de la nature des résultats du domaine de recherche. Il s'opère également une ouverture vers plus de transparence scientifique, plus de coopération entre équipes à travers le monde. Enfin, on voit apparaître de nouveaux paradigmes de recherche et on commence à envisager de nouvelles applications pratiques qui seraient difficiles à envisager sans ces corpus.

Philippe CARON

« *Restituer un dialecte phonétique du français vers 1700* »

La question du phonétisme et de la prosodie du français vers 1700 est devenue un enjeu bien après que des musiciens avant-gardistes en leur temps se sont efforcés de retrouver en musique instrumentale le timbre et la technique des instruments de l'époque dite baroque. Cette recherche appelait en toute logique un effort similaire dans les arts de la voix et Eugène Green a effectué le premier quelques travaux sourciers dans les années 1990 en ce sens, démontrant la pertinence d'une tentative.

Menée en partenariat avec le CRISCO de Caen, cette recherche vise à restituer d'abord et avant tout un artefact phonétique complexe qu'on peut appeler la diction haute du français vers 1700. Restituer, car tous les locuteurs sont morts et la tradition en est perdue.

La question des sources, de leur interprétation et des lacunes (malgré une période particulièrement abondante en données variées) sera abordée sur un cas d'école : l'oralisation des infinitifs en *-er* qui engage à la fois des questions de registre, de milieux sociaux, tout cela sur un nœud de variables sensibles à la liaison. On insistera enfin sur le fer de lance du projet qui est d'offrir une aide logicielle à la juste oralisation pour les acteurs et chanteurs à l'aide du synthétiseur de parole ©KALI.

TABLE RONDE

Olivier BAUDE

DGLFLF-Ministère de la Culture & LLL-EA3850-Université d'Orléans

« *Le programme Corpus de la parole* »

Le programme *Corpus de la parole* initié par la Délégation générale à la langue française et aux langues de France en partenariat avec le CNRS a pour vocation la constitution, la conservation, l'exploitation et la diffusion de corpus oraux constitués par les linguistes. Les nouvelles technologies de traitement des données linguistiques – des outils de transcriptions synchronisées et d'annotations aux normes de catalogage et de codage en passant par l'élaboration de base de données – ont un impact non négligeable sur les pratiques des chercheurs et la construction de leur objet quand il s'agit de répondre à un objectif de mutualisation et d'interopérabilité. Un premier bilan de ce programme offre l'occasion de présenter les problèmes rencontrés et les solutions proposées.

Antonia CRISTINOI

« *Traduction vs traduction automatique – un conflit insoluble ?* »

Le monde actuel est un monde dans lequel on constate une évolution constante des savoirs et des technologies, dont la transmission rapide est devenue indispensable. La traduction joue ainsi un rôle de plus en plus important, ce qui a comme effet la création d'un nombre croissant d'outils (traducteurs automatiques, outils d'aide à la traduction, etc.) destinés à faire face aux nombreuses demandes du marché. Si la création d'outils de TAO (traduction assistée par ordinateur) est généralement bien accueillie par les traducteurs, la TA (traduction automatique) est en revanche pour eux soit une source d'inquiétude soit un objet de mépris.

Dans ce contexte, il devient donc indispensable de montrer non seulement quelle place la TA peut occuper dans le travail du traducteur et sur le marché de la traduction, mais aussi de quelle manière traducteurs humains et linguistes peuvent contribuer à l'améliorer.

Jean-François SABLAYROLLES et Emmanuel CARTIER

« Une base de données pour la néologie : problématiques et solutions »

L'équipe néologie du laboratoire LDI a eu besoin d'une base de données non seulement pour engranger ses « prises » dans sa tâche de veille néologique, mais aussi pour analyser les néologismes relevés et l'utiliser comme moyen d'étude de la néologie du français contemporain. La base Neologia a d'emblée été conçue comme un véritable outil de travail. Une collaboration étroite et régulière entre les lexicologues et le linguiste-informaticien a permis de construire cette base qui répond aux objectifs fixés. Après une période de rodage d'un peu plus d'un an, la base a été corrigée en fonction des lacunes que son utilisation a permis de mettre au jour. Ses principes et son utilisation seront brièvement présentés.

Le pouvoir heuristique de la base Neologia sur notre travail n'a pas été négligeable. Les discussions lors de la création de la base ont conduit à approfondir la réflexion théorique dans le domaine en nous obligeant à expliciter jusque dans le détail nos objectifs, et son utilisation a permis de découvrir des problèmes auxquels on n'avait pas songé initialement. Ce qui a également contribué à des approfondissements théoriques ou à des précisions rassemblés dans un recueil à usage interne de l'équipe, mais qui ont fait et feront l'objet de communications et articles.

Michel SANTACROCE

« Internet et les nouvelles linguistiques : le « laboratoire Multimédia » et l'histoire des SDL »

Au-delà de l'aventure éditoriale particulière de feu-*Marges Linguistiques* qui sera évoquée (1999-2006/2007), je m'attacherai à mettre en évidence un point particulier qui s'articule autour d'un double mouvement contraire : d'une part une immense liberté fraîchement (re)conquise (les années 90^s) par certains linguistes grâce aux développements fulgurants des réseaux numériques ; d'autre part la reprise en main progressive des nouveaux domaines SDL en éclosion par des institutions et structures qui entendent récupérer pleinement le contrôle de la production et de l'agencement des connaissances linguistiques. La thèse défendue ici porte essentiellement sur le fait qu'aujourd'hui la naissance et la disparition d'un champ disciplinaire, d'un domaine ou sous-domaine de connaissances en SDL peut être appréhendé en un coup d'œil, en quelques années seulement. Cette disposition à l'accélération via les réseaux n'est en rien anodine, car elle permet une vision épistémologique « projective » et « rétroactive » qui n'a pas d'antécédent historique.

